A Design for a World Wide Web Decision-Support System using a Controlled Medical Terminology

Daniel Karlsson, Mats Carlsson, Urban Forsum, Ove Wigertz, Department of Medical Informatics, Linköping University and Department of Clinical Microbiology, University Hospital of Linköping, Sweden

We have developed a medical decision-support system based on Arden Syntax for Medical Logic Modules and hypertext using the World Wide Web. The two representations are integrated to provide a better platform for decision support. To manage the integration a controlled medical terminology has been used as a well-defined interface between the representations. The terminology is also used to facilitate communication between specialists and non-specialists.

INTRODUCTION

To support geographical distribution of expert knowledge among health care providers we have developed a Decision-Support System (DSS) using World Wide Web technique [1]. The system uses both Arden Syntax and hypertext for representing domain-specific medical knowledge. The Arden Syntax rules and the hypertext documents are integrated to provide explanations to given advice and to provide efficient information retrieval.

To facilitate the communication between the non-specialists, i.e. the users of the system, and specialist consultants, a Controlled Medical Terminology (CMT) is added to the system. The CMT is based on standardized nomenclatures and classifications, for example MeSH, IUPAC/IFCC[1] and ICD-10, to promote a uniform use of medical terminology.

In our design, we also observed the importance of using a CMT for organizing the knowledge base. Since our system is partly based on a large-volume hypertext knowledge base, there is an obvious need for a CMT to index the text masses, but we also recognized a need for a CMT to handle the connection between the hypertext and the MLM knowledge bases as well as the organization of the MLMs.

To enhance the functionality of the CMT a Data Dictionary (DD) is added to the system. The DD is used to keep information about what data is stored in which clinical databases and how the data is stored.

To examine the properties of our design we are developing a prototype in the field of Urinary Tract Infections (UTI) using a CMT based on the IUPAC/IFCC syntax and nomenclature [3]. The aim is to evaluate the system regarding usability in the general practitioners office.

BACKGROUND

Arden Syntax and Events

The Arden Syntax is a language for representing procedural medical knowledge in modules, so called Medical Logic Modules (MLMs) [4]. The MLMs consist of three categories: Maintenance, Library and Knowledge. The categories are built up of different slots. The first two categories mainly contain slots used for documenting the piece of knowledge represented in the MLM. In the knowledge category the semantics of the MLM is defined. The knowledge category contains four important slots: (1) the data slot, defining parameters to be used by the MLM, (2) the logic slot, defining the rule or rules of the MLM, (3) the action slot, defining the action to be performed if a rule in the logic slot is true, and (4) the evoke slot, defining the situation in which the MLM should be executed, i.e. which events trigger the MLM.

An MLM-controller [5] is used to schedule the execution of MLMs. This MLM-controller accepts events as input and responds by executing the MLMs which are triggered by the events at hand. Execution of MLMs can either be momentary, delayed and/or periodical. In a data-driven system an event could be the storage of critical patient data which requires checking. In a consulting system, such as ours, the events represents specific questions. An example of a question may be what treatment to use for a specific disease and a specific patient.

Hypertext

Hypertext is a technique which relaxes the constraints which applies to normal, printed text [6]. Ordinary text is organized in a linear manner. Pages of a book or an article follow each other

[1] International Union of Pure and Applied Chemistry/ International Federation of Clinical Chemistry

linearly. In hypertext the pages are instead organized according to semantic content. This means that hypertext pages of similar or otherwise related content are linked together. These links can either be embedded in the hypertext pages or be defined in a independent semantic network [7]. An independent semantic network can consist of the concepts and interrelationships of a CMT [8].

One of the most spread hypertext systems is the World Wide Web (WWW). WWW is a client/server system for distributing hypertext documents over the Internet or other networks. The hypertext documents are represented in the HyperText Markup Language (HTML). HTML provides means of authoring and storing multimedia documents with hypertext links according to the embedded model. The pages may be dynamically created by programs that conform to the Common Gateway Interface (CGI) standard.

Original System Design

The original idea behind the design of our system is to integrate hypertext and Arden Syntax representations and thereby provide a better platform for decision support to health care providers [1]. A client/server architecture has been used for the system. The clients are regular HTML browsers which are connected to a WWW server through a TCP/IP network. The WWW server is connected to hypertext and MLM knowledge bases via ordinary HTML links and CGI calls.

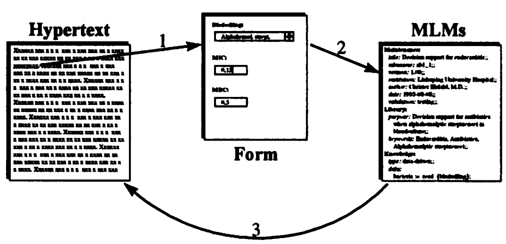

Figure 1 - Integration

The two knowledge bases are integrated using hypertext links and Arden Syntax events (see Figure 1). Anchors in the hypertext documents are linked to HTML forms (1) used for patient data entry. The form has associated Arden Syntax events (2) included as invisible input elements. The event defines a set of MLMs where all MLMs concern the same specific question. The MLMs have hypertext links in the action slot to hypertext documents (3) which gives the background knowledge needed for understanding the advice produced by the MLMs.

The Controlled Medical Terminology

SpriTerm is a data model for a CMT developed using the Microsoft ACCESS™ personal database management system. Its purpose is to supply the Swedish health care community with terms with unambiguous meaning in order to assist communication between health care actors and health care systems [9]. SpriTerm is the data model used in this project.

The CMT consists of terms and concepts connected in a semantic network. Concepts can have several terms connected to themselves. If a term is associated with the same concept as another term it means that they are synonyms. This can be used to cover both professionally defined and trivial terms. Every concept has relations to other concepts. This means the users can traverse the semantic network, looking for the concept he is interested in. Other information is also linked to the concepts, such as coding and classification systems, for example ICD-10, SNOMED III and so on.

One of the central relations is the specialization relation: ISA. This relation enables saying that *Blood* ISA *Body fluid*, which means that *Blood* has all the properties of the concept *Body fluid*. The *Blood* concept then has some characteristics distinguishing it from its parent, such as *Red blood cell count*.

Even if the ISA-relation is pivotal due to its role in the inheritance process it is far from the only relation in the semantic network. In fact, any other relation wanted could be created. One of the features of this CMT is that it treats all relations, except ISA, as concepts. This way they also can benefit from the inheritance mechanism. Other relations which are often used are: PartOf, HasValue, HasLocation.

METHOD

The Data Dictionary

In Huff et. al. [2] a Medical Data Dictionary (MDD) is described as a system consisting of a vocabulary, i.e. a CMT or thesaurus, an information model and a knowledge base. The vocabulary concerns organization of medical terms and concepts, the information model describes the connection between vocabulary items and clinical databases, and the knowledge base includes relationships and rules which support clinical decisions. Our view of a DD covers this information model.

The DD is used to define the correspondence between concepts in the CMT and data as it appears in various physical databases. The concept represents the abstract idea of an object. Instances of concepts may have values. For example, instances of descendants of the concept *Lab result* do have values whereas instances of the concept *Bone* do not. The value of an instance of a concept is stored as data in a database. Furthermore the data corresponding to a concept must have a type to determine the set of allowed values of the data. This type is the same for all data corresponding to the same concept.

The information in the DD is used to provide information concerning concept to data mapping:

- In which databases can data be found which corresponds to certain concepts.
- How can the data be retrieved from a database.
- What are the type, unit, format, and the set of allowed values for a datum corresponding to a certain concept.

A way to handle access to databases is discussed in Johansson et. al. [10]. To retrieve data for parameters in the data slot of MLMs, SQL-templates for different Arden Syntax data access methods are used. The templates are instantiated with information from the data slots of the MLMs and information from the CMT and the DD.

The type of data for a concept may be a simple data type such as integer, real, or string. An examples of such a concept is *Concentration (QU50043)* as measured in number of bacteria per liter. The type of data may also be an enumerable type, i.e. a nominal or ordinal scale, for example the scale {S, I, R} for sensitivity measurement. The possible values for enumerable types is specified in the DD.

A third kind of type is needed when an enumerable type is not sufficient. An example is the set of allowed values of data corresponding to the concept *Urine culture (NPU06073)* as defined by IUPAC/IFCC. One approach would be to enumerate all possible bacteria which could appear in a urine sample. But in that case the level of granularity in describing the bacteria must be defined statically at the time for construction of the DD. This can be a problem since the level of granularity can not be adapted to suit different situations and different user's needs. For example, a health care provider without specialist knowledge in microbiology may discuss bacteria on the genus level while a specialist refers to bacteria on the species or even the biotype level. There is a need to specify the value of a concept using different levels of granularity in different situations. This can be accomplished by letting the type of data defined in the DD refer to the CMT. For example (see Figure 2), the type of values of the concept *Urine culture*, as defined in the DD, is the concept *Bacterium*.

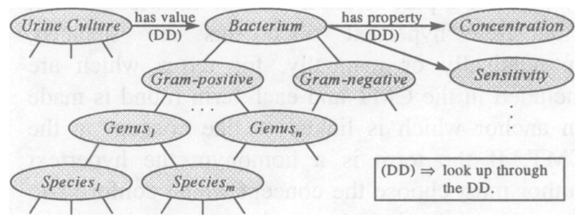

Figure 2 - A complex data type

The concept *Bacterium* is also defined in the CMT and has descendants which describe specializations of the concept, all with professional definitions. When assigning a value to data corresponding to the concept *Urine culture* valid choices are the descendants of the concept *Bacterium*. This kind of type is called a complex data type. The concept *Bacterium* also has values; for example, the concentration of the bacterium in the sample and its sensitivity to antibiotics.

System Design

The use of hard-coded, or embedded, links, as in the original design, is not without problems. When the size of the hypertext knowledge base grows the number of potential links increases rapidly. Every time a new hypertext document is added to the knowledge base every previous document is a candidate for linking, either to or from the new document. Every time a hypertext document is deleted all other documents must be examined. When there are many hypertext authors working together on one knowledge base their work must be coordinated, at a high cost, so that no relevant links are neglected. Most likely, especially when working with continuous updating on large knowledge bases, relevant links will be forgotten resulting in a low recall on hypertext documents. More about problems concerning hypertext systems can be found in Conklin [6] and Cimino et. al. [8].

Our system design is based on a CMT together with a DD, functioning as an interface between the hypertext and the MLM knowledge bases.

The system design allows four main functionalities: (1) to organize hypertext using the CMT [8], (2) to generate forms for patient data entry based on the CMT and the DD, (3) to organize the connection between hypertext documents and MLMs, and (4) to provide explanations to advice produced by the

MLMs, this by searching the hypertext knowledge base using concepts in the CMT.

When organizing the hypertext documents according to a CMT the hypertext documents are first indexed using terms from the CMT. The index is stored in an inverted index database (see Figure 3) which is a part of the hypertext knowledge base. Then the hypertext documents are scanned, automatically or manually, for terms which are included in the CMT and each term found is made an anchor which is linked to one concept in the CMT. If the term is a homonym the hypertext author must choose the concept which conforms to the writers intention.

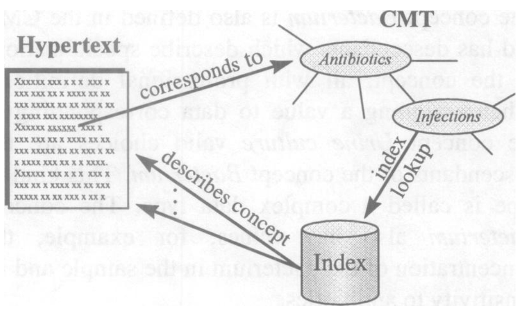

Figure 3 - Hypertext and the CMT

When a user chooses a hypertext link he/she is presented with information about the corresponding concept, including professional definitions, codes, recommended and alternative terms together with the list of hypertext documents which relate to the concept as stated in the index database.

The second functionality offered by the design is automatic or semi-automatic construction of forms. A form is a collection of User Interface Objects (UIOs) used to enter patient specific information for use by the MLMs. Examples of UIOs are menus, buttons and text fields. Each UIO in a form corresponds to a parameter that is needed by the MLMs to give advice regarding a specific question. One form may be used for one or many specific questions. The questions are defined by Arden Syntax events (see above). Consequently, one form includes one or many Arden Syntax events. A form must consist of all parameters used in the data slots of the MLMs which are triggered by the events included in the form. Some of the parameters used are retrieved from clinical databases such as laboratory databases or patient records and can not be altered.

The UIOs are constructed based on the concept in the CMT which corresponds to the parameter using type information from the DD. For concepts with numerical values the UIO becomes a numerical input field, for concepts with an enumerable value type the UIO becomes a menu or a check box etc. For complex data types (see above) more complex UIOs are needed. The form could be generated either from a list of concepts provided by the MLM author or generated automatically from the data slots of MLMs.

The third functionality offered by the system is to organize the integration of hypertext documents and MLMs (see Figure 4). A central structure in the integration is the Event Information Table (EIT). The EIT links concepts in the CMT with Arden Syntax events. A concept may have many associated Arden Syntax events corresponding to various specific questions related to the concept. When a user considers a concept in the CMT a, possibly empty, list of questions is given. For example, when considering the concept *Antibiotics* the question "Which antibiotic shall I use for my patient?" is presented. When the question is selected a form corresponding to the event is presented. After entering patient data the event is sent to the MLM controller. The MLM controller executes the corresponding MLMs and gives advice back to the user.

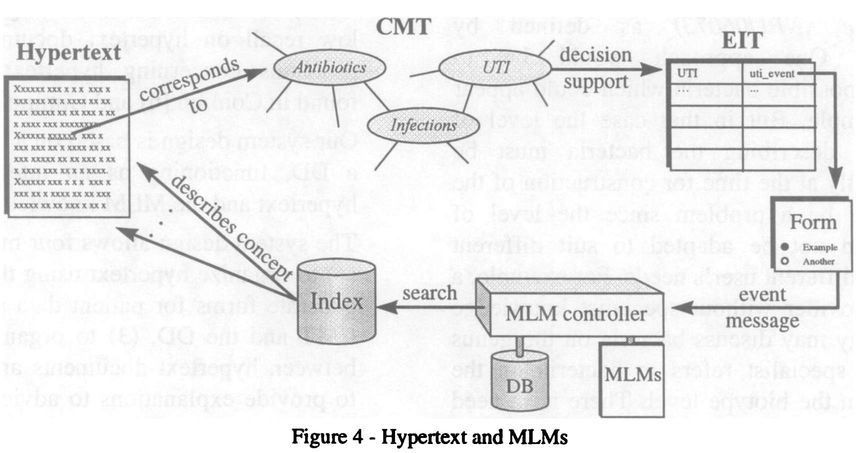

Figure 4 - Hypertext and MLMs

The fourth functionality offered is the possibility to give explanations to given advice by searching the index for suitable hypertext documents (see Figure 4). Every advice produced by the MLMs include a query to the index database specific to the advice. The system justifies the advice produced by the MLMs by providing links to hypertext documents which support and give background knowledge to the standpoint of the system. By disallowing hard-coded links from MLMs to hypertext documents and instead using the index database, maintenance is made easier and recall is improved, but the precision may suffer. To keep precision at an acceptable level a sufficiently strong query language must be used. Various techniques, for example probabilistic retrieval and fuzzy sets, are discussed in Salton [11].

A method to further enhance the usability of the system is to support individualization of the system to reflect the individual user's preferences. When searching for hypertext documents in the index database the user's staff category, background knowledge, and current interest can be considered.

DISCUSSION

With this design the domain experts may use both rules and multimedia documents to express their knowledge. The construction of such a knowledge base is a hard task if the system does not provide support that frees the expert from managing the complexity of the integration.

There are two main advantages of using a CMT together with a clinical DSS. First, the CMT can be used to promote the use of professionally defined terms. Second, the CMT facilitates the maintenance of the knowledge base.

A problem may arise when the domain includes vast amounts of relations between concepts. The concepts may be hard to survey and navigating among concepts in the CMT may become too complex. Therefore it is important to filter the CMT to include only the subset of concepts which are relevant to the current domain. There should also be possibilities in the user interface to reduce the complexity of the CMT by, for example, excluding certain relations.

We believe that this design brings enhanced availability of specialist knowledge suited for the user's needs. Furthermore, we believe that this design facilitates knowledge authoring and organization of the knowledge base.

Acknowledgments

This work was supported by NUTEK, the Swedish National Board for Industrial and Technical Development (Project no. 94-4876P).

References

1. Karlsson D, Ekdahl C, Wigertz O, Shahsavar N, Gill H, Forsum U. Extended Telemedical Consultation Using Arden Syntax Based Decision Support, Hypertext, and WWW technique. Meth Inform Med. Submitted.
2. Huff SM, Cimino JJ. Medical Data Dictionaries and their Use in Medical Information System Development. In: Prokosch HU, Dudeck J, eds. *Hospital Information Systems: Design and Development Characteristics; Impact and Future Architecture.* Amsterdam: Elsevier 1995.
3. Olesen H. Properties and Units in the Clinical Laboratory Sciences, I. Syntax and Semantic Rules. Eur J Clin Chem Clin Biochem 1995; 33:627–636.
4. Hripcsak G, ed. Special Issue: The Arden Syntax. Comp Biol Med 1994; 5.
5. Gao X, Johansson B, Shahsavar N, Arkad K, Åhlfeldt H, Wigertz OB. Pre-compiling Medical Logic Modules into C++ in Building Medical Decision Support Systems. Comp Meth Prog Biomed 1993; 2: 107–119.
6. Conklin J. Hypertext: An Introduction and Survey. IEEE Computer 1987; 17: 17–41.
7. Rada R. *HYPERTEXT: from Text to Expertext.* London: McGraw-Hill 1991.
8. Cimino JJ, Elkin OL, Barnett GO. As We May Think: The Concept Space and Medical Hypertext. Comp Biomed Research 1992; 3: 238–263.
9. Carlsson M, Wigertz OB, Junfors A, Wallin SB. Development of a terminology management system for the Swedish healthcare sector. In: *MIE 96.* Accepted.
10. Johansson B, Shahsavar N, Åhlfeldt H. Data Mapping Method for Arden Syntax Knowledge Modules. Meth Info Med. Submitted.
11. Salton G. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Reading, Mass.: Addison-Wesley 1989.